DOCUMENT RESUME

ED 279 706 TM 870 093

AUTHOR Williams, Paul L.
TITLE NAEP Writing Assessment: A Committee Member's Perspective.
PUB DATE Sep 86
NOTE 4p.; One of 46 papers commissioned by the Study Group on the National Assessment of Student Achievement and cited in Appendix B to their final report "The Nation's Report Card" (TM 870 049). For other papers in this group, see TM 870 050-094.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; Educational Testing; Elementary Secondary Education; Evaluation Criteria; Holistic Evaluation; *Measurement Objectives; *National Surveys; *Scoring; State Programs; State Surveys; *Testing Programs; *Writing Evaluation
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

A major issue facing the National Assessment of Educational Progress (NAEP) Writing Assessment is its relationship with the most frequently encountered measurement practices among State Education Agencies (SEAs) and Local Education Agencies (LEAs). Should NAEP conduct its assessments in a fashion that represents how things should be, or reflect the status of how things really are in the nation's schools? Clearly, at some point the debate must be resolved so that program planning can move forward. The predominant method of scoring writing within NAEP is Primary Trait. Most state and local assessments do not use trait scoring. States with well developed writing programs use either focused holistic or analytic scoring, or a combination of the two. NAEP scoring might better reflect an appropriate practice by addressing factors such as: (1) the identification of elements of composition that are desirable given current practice and curricular/instructional direction among the state and local education agencies; (2) the desirable mix of text level and sentence level elements; (3) the applicability of the scoring criteria at the classroom level; (4) the applicability of the scoring criteria across grades; (5) the level of competence elicited by the criteria; and (6) the composition of the group that determines the scoring criteria and rater quality. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED279706

NAEP Writing Assessment

A Committee Member's Perspective

Paul L. Williams
CTB/McGraw-Hill

September, 1986

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. L. Williams

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper commissioned by

THE STUDY GROUP ON THE NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENT

1986

TM 870 073

BEST COPY AVAILABLE

NAEP WRITING---SCORING AND ADMINISTRATION

One of the major issues facing the future development of the NAEP Writing Assessment, as well as the structure of the program as a whole, is its relationship with the most frequently encountered measurement practices among State Education Agencies (SEAs) and Local Education Agencies (LEAs). On the one hand, the case could be made that NAEP should lead the assessment community, and conduct its assessments in a fashion that represents how things should be. Conversely, the case might be made that NAEP should reflect the status of how things really are in the nation's schools.

The debate surrounding the proper place for a national assessment on the continuum representing how things should be and how things actually are could extend indefinitely. The central issue, however, is that at some point the debate must be resolved so that program planning can move forward. Thus, the purposes of the assessment must be clearly articulated, for it is the purposes of the assessment which must ultimately guide its design and implementation.

In the areas of the scoring and administering writing assessments there are a number of areas that might be explored which would enhance the position of the NAEP assessment relative to current practice.

SCORING METHOD

The predominent method of scoring writing within NAEP is Primary Trait. Perhaps the most noteworthy of the states which use trait scoring in some form are California, Indiana, Wisconsin, and Louisiana. California and Wisconsin have provision for holistic or analytic scoring as well.

Most state and local assessments do not use trait scoring. States with well developed writing programs such as Maryland, South Carolina, Georgia, Texas, New Jersey, New York, North Carolina, Oklahoma, Nevada, Connecticutt, Illinois, Oregon, Mississippi, Alabama, Hawaii, Idaho, Maine, and Massachusetts use other than Primary Trait scoring as their primary method. The scoring may be either focused holistic, analytic, or a combination of the two.

NAEP scoring might better reflect an appropriate practice by addressing factors such as:

1. the identification of elements of composition that are desirable given current practice and curricular/instructional direction among the state and local education agencies.
2. the desirable mix of text level and sentence level elements.
3. the applicability of the scoring criteria at the classroom level.
4. the applicability of the scoring criteria across grades.
5. the level of competence elicited by the criteria, and
6. the composition of the group that dertermines the scoring criteria.

Point #6 above is included because of the lack of participation in objective or scoring guide development on the part of state and local district personnel. As the purposes of NAEP may be expanded in the future it is imperative that those who are effected by the scores have a substantive opportunity to participate in the design of the system they may be using.

The need for this participation is particularly clear in the development of the holistic scale. In the NAEP assessment table leaders and scoring directors used a procedure associated with classical scoring, where the perceived need for a set of scores to approximate a normal distribution overshadowed instructional and curricular considerations. This approach is not acceptable practice in any state program of which the committe is aware.

Further, since there is little articulation among the levels assessed by NAEP in the characteristics found at score points, perturbations will manifest themselves when growth among age/garde levels is compared. For example, those across the nation who are interested in generalizing growth patterns may inappropriately assume that a 5% increase in students at the score range of 4-6 for age 17 represents the same amount of skill acquisition as a 5% gain at age 9. Such apparent movement may be as attributable to the characteristics of the scale as real growth.

RATER QUALITY

Rater quality is the ultimate test of the quality of the scores that energe from the assessment. There is insufficient substantiation of rater quality indices found in the NAEP writing documents. For example, prevalent practice in leading writing programs ensures that after training a rigorous qualification process is followed that guarantess, to as large an extent possible, the best raters move forward into scoring student papers. The NAEP program should be able to guarantee that their raters are at least as qualified as those in most of the leading progams in the states and local districts.

Further documentation, and where appropriate, additional practices need to be adopted to ensure that rater quality during scoring is as good as current practice in other large assessments. Providing a 20% sample of double scored papers does little to enhance the credibility of the quality of ratings. Additionally, data needs to be provided on how well the raters are able to remain on the scoring guide throughout the scoring sessions. This validity check is crucial because teams, or even rooms, of raters may systematically drift off the criteria. In so doing they may still have what appear to be acceptable consistency indices, yet the criteria may be mis-applied.